Yangon University of Economics
Post Graduate Diploma in Research Studies (PGDRS) Progamme
Second Quarter Examination, December 2023

PGDRS-211
PGDRS (9th Batch) **Regression Analysis**
Answer **All** Questions. **Time Allowed: 3 hours**

1. Determine whether each statement is **True** or **False**.
   - (a) The standard error of estimate is a measure of the average amount by which the actual observations for vary around the regression line.
   - (b) The explanatory variable is called dependent variable.
   - (c) Regression determines the magnitude of the change in Y given change in X.
   - (d) Multiple regression involves two or more dependent variables.
   - (e) If the relationship can be described by a curve, linear regression is used.
   - (f) Multicollinearity exists if one of the independent variables is linearly related to any of the dependent variables.
   - (g) If r > 0, b1 will be negative and the line will slope up.
   - (h) In the presence of heteroscedasticity, the regression coefficients become more efficient.
   - (i) If heteroscedasticity is suspected, the use of the generalized least-squares (GLS) method is recommended.
   - (j) If the variance in errors is not the same for all values of X, heteroscedasticity occurs.
   - (k) $nR^2 > \mathcal{X}^2_{\alpha,k}$ $means\ that$ error variances are equal.
   - (l) Time-series data are a collection of data over a series of time periods (days, months, year,etc.)
   - (m) Variance inflation factor is used to detect correlation among independent variables.
   - (n) The error is the difference between the actual value of the dependent variable and the estimated value of the dependent variable.

2. Hutton was interested in the relationship between a person's income and the amount of money they had invested in the stock market. Fifty individuals were randomly selected on the presumption that an individual's investments in the stock market are influenced by his or her income. Using regression and correlation analysis, do these data suggest such a relationship? (Data are in thousands of dollars.)

   $\sum X = 9,385$ $\sum X^2 = 3,025,553$ $\sum Y = 988.1$ $\sum Y^2 = 32,224.51$ $\sum XY = 303,471.3$
   - (a) Identify the dependent and independent variables.
   - (b) Calculate and interpret the results of the regression model.
   - (c) Are the data cross-sectional or time-series?

3. Owners of Mighty Muscles, Inc., a health club in Hoboken, New Jersey, are trying to increase membership at the club. After reviewing operations at other health clubs in the area, Harry Hunk, the manager at Mighty Muscles, feels that there might be a relationship between the number of members in a health club and several explanatory variables, including monthly dues, local population, and type of weight training equipment in the club. Three kinds of weight machines are found in health clubs: MuscleUp (MU), Iron Man (IM), and Studly Dude (SD). Mr. Hunk collects data for these variables for 20 health clubs similar to Mighty Muscles.

Membership	**Dues**	**Population (in 1,000's)**	**Equipment Type**
345	$22.50	39.7	MU
387	21.00	43.3	MU
412	20.00	47.3	MU
434	20.00	48.9	SD
467	18.50	54.5	SD
490	18.00	55.7	SD
612	15.00	62.1	SD
523	17.50	56.7	IM
545	17.00	58.9	IM
567	16.50	60.0	IM
578	16.00	61.2	SD
599	15.00	63.6	SD
634	14.50	60.0	IM
656	13.00	61.0	SD
669	13.10	61.4	MU
134	30.00	23.5	MU
243	27.00	27.8	IM
259	25.50	32.1	IM
278	23.00	37.3	SD
321	24.00	35.8	SD

(a) Construct the data set with the proper dummy variables. Indicate the coding system used.
(b) Solve the regression model and interpret the results.
(c) Test the regression model at 5% level.
(d) Test each regression coefficient at 5% level.
(e) Is there evidence of multicollinearity between monthly dues and local population?
(f) Compare your results in the previous problem based on VIF.
(g) Which variable is excluded in the model?

4. Data on 20 cars about MPG (average miles per gallons), HP (engine horsepower), SP (top seed, miles per hour), and WT (vehicle weight in 100 lbs.).

Observation	**MPG**	**SP**	**HP**	**WT**
1	65.4	96	49	17.5
2	56.0	97	55	20.0
3	59.2	98	62	22.5
4	53.3	98	62	22.5
5	42.2	109	90	25.0
6	40.9	110	92	25.0
7	46.9	90	52	27.5
8	36.3	112	103	27.5
9	36.1	103	84	27.5
10	33.2	109	102	30.0
11	32.9	109	102	30.0
12	32.3	120	130	30.0
13	33.7	109	115	35.0
14	32.6	109	115	35.0
15	31.3	109	115	35.0
16	25.6	107	120	40.0

17	25.3	114	140	40.0
18	23.9	114	140	40.0
19	22.9	110	140	45.0
20	22.9	110	140	45.0

(a) Estimate the parameters of this model and interpret the results.
(b) Would you expect the error variance in the model to be heteroscedastic? Why?
(c) Use the White test to find out if the error variance is heteroscedastic.